

DOCUMENT RESUME

ED 189 153

TM 800 323

AUTHOR Tate, Richard L.
TITLE Multivariate ATI Analysis.
PUB DATE Apr 80
NOTE 36p.: Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Aptitude Treatment Interaction; Discriminant Analysis; *Hypothesis Testing; Individualized Instruction; Learning Modules; *Multivariate Analysis; Reading Ability; School Attitudes; Science Education; Secondary Education; Statistical Significance; Time Factors (Learning)
IDENTIFIERS Individualized Science Instructional System

ABSTRACT

This paper reviews multivariate hypothesis testing, interval estimation, and discriminant analysis procedures for an aptitude-treatment interaction (ATI) analysis and illustrates the procedures with data from a large experimental study of science education at the secondary school level. The treatment variable of interest is time allowed for study with three controlled levels and the student aptitude is reading ability. The treatment and aptitude variables are shown to interact for four dependent variables representing achievement and attitude, and the interaction is described. The multivariate linear model was used for the analysis. Two advantages of the model are discussed. The inferential payoff is better control of Type I errors in hypothesis testing and confidence levels for interval estimation; and the potential for simplification of the ATI description is greater. (Author/CTM)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED189153

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Multivariate ATI Analysis

Richard L. Tate

Florida State University

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

R. Tate

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

TM 800323

Running head: Multivariate ATI

Abstract

A multivariate aptitude-treatment-interaction (ATI) analysis is demonstrated for data from a large science education study. The interaction between student reading ability and the time allowed in an individualized instructional approach is described for multiple achievement and attitudinal outcome variables. The discussion includes a review of the rationale for use of multivariate analysis and interpretations of the hypothesis testing and interval estimation results.

Multivariate ATI Analysis

Most studies which have included multiple outcome variables in aptitude-treatment-interaction (ATI) research (e.g., Peterson, 1977) have considered each outcome variable separately in analyzing the data. As the number of outcome variables increases, the usual reasons for a multivariate analysis become pertinent: control of error rate in testing (or control of confidence level for interval estimation) and parsimonious description of the ATI with optimal combinations of the outcome or dependent variables. Several authors have considered aspects of the ATI multivariate generalization. For example, Potthoff (1964) discussed simultaneous interval estimation for multiple dependent variables, Lo (1973) provided an illustration of the test of the global multivariate ATI hypothesis, and Timm (1975, p. 340) considered the same global hypothesis in the context of testing the homogeneity of regression assumption for multivariate analysis of covariance. There appear to be no published studies, however, which illustrate simultaneously all major facets of a multivariate ATI analysis.

This paper reviews multivariate hypothesis testing, interval estimation, and discriminant analysis procedures for an ATI analysis and illustrates the procedures with data from a large experimental study of science education at the secondary school level. The treatment variable of interest is time allowed for study with three controlled levels and the student aptitude is reading ability. The treatment and aptitude variables are shown to interact for four dependent variables representing achievement and attitude, and the interaction is described.

Example Data

The analysis is illustrated with results from a recent field test of the Individualized Science Instruction System (ISIS), an instructional package of life and physical science topics for high school students used nationally since 1974 (Burkman, 1974). The field test considered several management approaches for a complete series of instructional modules used throughout the 1978-79 school year in various parts of the country. The data used herein is for an individualized management approach using a set of three life science modules covered at the beginning of the school year. Each student independently studied module materials, did exercises, took self-tests, and "recycled" through the materials if necessary. Each student was responsible for pacing him or herself in order to complete the materials for the three modules in the allowed time specified by the teacher. There were 413 students in the analysis sample for this illustration; the unit of analysis was assumed to be the individual student.

Allowed time was a variable of major interest in the field test because of its obvious role in the possible trade-off of the degree of student mastery of each topic covered and the number of topics covered during a school year. Participating teachers were randomly assigned to one of three levels of allowed time (Low Time or LT = 2750 minutes, Medium Time or MT = 3000 minutes, and High Time or HT = 3250 minutes). For a 50-minute class period, these allowed time conditions corresponded to 55, 60, and 65 class periods, respectively, for all three modules.

The four dependent variables in the analysis were measured at the completion of the allowed time for the three modules. The two achievement variables represented percent of items correct for a set of items covering "core" material (core achievement or ACHC) and percent correct for advanced material studied after completion of the core material (advanced achievement or ACHA). The two attitudinal outcome variables were the number of science courses which a student planned to take before graduation (NCRSE) and the student's general interest in science on a scale from 1 to 4 (INTRST). The student aptitude of interest in the aptitude-treatment-interaction was reading ability measured with a test administered to all students at the beginning of the school year.

This description of the study conditions and data has been brief because the data are used only for illustration of an analysis technique. Substantive interpretations of the results would have to take into consideration several complications common in field research, e.g., differential attrition among the treatment levels and lack of independence among individual responses due to individuals working with the materials in the same class with other study students.

ATI Model

The multivariate linear model (see, e.g., Timm, 1975) is used for the analysis. Thus, if Y_1' is a p-variate vector observation from a multivariate

normal distribution with variance-covariance matrix Σ , the expected value vector $E(Y_1') = \mu_1'$ is modeled with

$$\mu_1' = \sum_j \beta_j' X_{1j} \quad (1)$$

where X_{1j} is the j^{th} independent variable ($j = 1, \dots, q$) for the i^{th} observation ($i = 1, \dots, N$), and β_j' is the $1 \times p$ vector of model parameters for the p dependent variables which is associated with the j^{th} independent variable. The coding used for the current example with one treatment factor

(allowed time) having three levels and one aptitude (reading ability) was:

$X_{11} = 1$, $X_{12} = \text{aptitude (A)}$, $X_{13} = 1$ for treatment level 1 and $= 0$ otherwise,

$X_{14} = 1$ for treatment level 2 and $= 0$ otherwise, $X_{15} = X_{12} \cdot X_{13}$, and

$X_{16} = X_{12} \cdot X_{14}$. Decomposition of the resulting full rank model by substitution of dummy coding results in the following models for each treatment level

$$\begin{aligned} \mu_1' &= (\beta_1' + \beta_3') + (\beta_2' + \beta_5')A \\ \mu_2' &= (\beta_1' + \beta_4') + (\beta_2' + \beta_6')A \\ \mu_3' &= \beta_1' + \beta_2'A \end{aligned} \quad (2)$$

where μ_i' now refers to the p -variate expected value vector for the i^{th} treatment level and a specific aptitude value A . As expected with the 1, 0 coding, the model parameters reflect control group comparisons. That is, β_1' and β_2' are the intercept and slope coefficient vectors for the last (control) treatment level, while β_3' and β_4' represent intercept differences and β_5' and β_6' slope differences between levels 1 and 2 and the control level. When β_5' and β_6' , the coefficient vectors associated with the aptitude-treatment-interaction, are both equal to the $1 \times p$ null vector, all regression lines for each dependent variable are parallel, i.e., there is no ATI.

The linear model stated in equation 1 has several classical variations. Graybill (1976) distinguishes in the univariate case between the "general linear model" in which the X's are assumed to be fixed and the "linear regression model" in which the X's are random. In the fixed model, the analysis sample is assumed to result from sampling $\underline{Y}$ vectors from the multivariate normal distributions associated with each combination of fixed X's of interest. On the other hand, the random model assumes random sampling of $[\underline{Y} \ \underline{X}]$ vectors from a joint multivariate distribution, and further assumes that the conditional distribution of $\underline{Y}$ given $\underline{X}$ is multivariate normal (Graybill's Case II, p. 375). The usual sampling in ATI research involves joint sampling of the Y's and the aptitudes from each of several populations (treatment levels). Thus, the model in the form of equation 1 is a mixed model with the fixed X's corresponding to the dummy variables used for group membership and random X's for aptitudes. The equivalent model in equation 2 consists of multiple random models. Graybill shows that even though the fixed and random models are fundamentally different, they result in identical computational equations for hypothesis testing (although power is different for the two models) and interval estimation. Therefore, the computer program by Tinn and Carlson (1973), which is based on the full-rank fixed model, was used for computations in the present case.

The parameter estimates for the current example are given in Table 1 and result in the interaction patterns shown in Figure 1. These are the same estimates and patterns that would have been obtained with multiple univariate analyses. Ignoring, for the moment, the question of statistical significance,

some modest but practically significant interactions are found in Figure 1. For the achievement variables, ACHC and ACHA, there is a general negative relationship between achievement and allowed time, but an ordinal interaction between allowed time and reading ability results in variation of the size of the allowed time effect with reading ability. For example, the superiority of the LT condition over the HT condition for ACHA decreases from approximately 24 percentage points at an ability of 30 to nine percentage points at an ability of 80. An estimated disordinal interaction is shown in Figure 1 for the attitudinal variables, NCRSE and INTRST. The LT condition is superior across the ability range for both variables, but there is a reversal of superiority associated with the MT and HT conditions.

Insert Table 1 and Figure 1 about here

Hypothesis Testing and Interval Estimation

There are two potential advantages in the use of a multivariate analysis rather than multiple univariate analyses. The inferential payoff is better control of Type I errors in hypothesis testing and confidence levels for interval estimation. A multivariate global null hypothesis can be expressed in the form $H_0: C_0' B D_0 = 0$ where the rows of the $q \times p$ matrix B are the vectors β_j' , the rows of the $df_h \times q$ matrix C_0 specify df_h linearly independent functions of the model parameters and the columns of the $p \times t$ matrix D_0 specify t linearly independent combinations of the original p variables. The null hypothesis can be tested with any of several test criteria based on the characteristic roots of $H E^{-1}$, where E is the residual sums of squares and cross products matrix associated with the unrestricted model and the hypothesis matrix H is the increment in the residual matrix

due to constraining the model with the null hypothesis (see, e.g., Timm, 1975, p. 208). The test criterion used here is λ_1 or the largest root of $\underline{H} \underline{E}^{-1}$. The null hypothesis is rejected if λ_1 is larger than $\lambda_1^\alpha(s, m, n) = [\theta^\alpha / (1 - \theta^\alpha)]$ where θ^α tables and charts are given in Timm (1975). The degree of freedom parameters are $s = \min(df_h, t)$, $m = (|t - df_h| - 1)/2$ and $n = (df_e - t - 1)/2$ where $df_e = N - q$.

Gabriel (1968) has discussed a class of simultaneous test procedures which can be used to follow a rejected global hypothesis with tests on any implied hypotheses, controlling the family-wise error rate. An implied hypothesis can be expressed as $H_1: \underline{C}_1' \underline{B} \underline{D}_1 = \underline{0}$ where each of the linearly independent rows of $\underline{C}_1'$ is a linear combination of the rows of $\underline{C}_0'$ and each of the linearly independent columns of $\underline{D}_1$ is a linear combination of the columns of $\underline{D}_0$. The computed multivariate test statistic for each implied H_1 is compared with the critical value of the test statistic for the global H_0 , λ_1^α in this example. Results from this procedure are coherent; that is, an implied hypothesis at some level of generality can only be rejected if the higher level implying hypothesis is rejected.

An index of multivariate strength of association which can be used for the global and implied hypotheses is $n_{LH}^2 = V/(1 + V)$ where V is the Lawley-Hotelling trace criterion, $\Sigma \lambda_1$. Stevens (1972) and Huberty (1972) have considered this and several other indices for the one-way MANOVA case, and have concluded that all indices produce similar results. In addition, the usual univariate results can also be recovered from the $\underline{H}$, $\underline{E}$, and $\underline{T}$ (total) matrices for each hypothesis. Thus, the ratio of a diagonal element

of $\underline{H}$ to the corresponding element of $\underline{E}$ is $SSH/SSE = \Delta R^2 / (1 - R_F^2)$ for a single dependent variable, where $\Delta R^2 = R_F^2 - R_R^2$, R_F^2 is the coefficient of determination for the full or unrestricted model, and R_R^2 is that for the model restricted by a true null hypothesis. The univariate F statistic is $[\Delta R^2 / df_h] / [(1 - R_F^2) / df_e]$.

The ultimate detail in the follow-up to a rejected global H_0 consists of tests or interval estimates for single scalar parametric functions called contrasts here. A contrast is defined as $\gamma = \underline{c}' \underline{B} \underline{d}$ where $\underline{c}'$ is a $1 \times q$ row vector and $\underline{d}$ is a $p \times 1$ column vector. Later descriptions are simplified if a model parameter for a specific combination of dependent variables (defined by $\underline{d}$) is represented by $\beta_j = \beta_j' \underline{d}$. Thus, $\gamma = \sum c_j \beta_j$ where the c_j are the q elements of $\underline{c}'$. The $(1 - \alpha)$ 100% confidence interval on γ is given by

$$CI(\gamma) = \hat{\gamma} \pm C S_{\hat{\gamma}}$$

where $\hat{\gamma} = \sum c_j \hat{\beta}_j$, $S_{\hat{\gamma}}$ is the estimated standard error of $\hat{\gamma}$, and the critical value C determines the nature of the confidence statement. A simple confidence interval results if $C = t^{\alpha/2}$ while simultaneous intervals are obtained with the use of the various critical values associated with the global hypothesis. Gabriel (1968) has shown that the narrowest intervals are obtained using the λ_1 criterion; thus, for this example, $C = \sqrt{df_e \lambda_1^\alpha}$. The interval estimate with this critical value was first proposed by Roy and Bose (1953) and results in control of the family-wise confidence level for

the infinite family of contrasts implied by the global H_0 . The result of the simultaneous test of the null hypothesis associated with a contrast is easily obtained by noting whether the hypothesized value lies within the interval.

The Bonferroni Inequality (Miller, 1966, p. 8) provides an alternative approach to the construction of simultaneous intervals for any preplanned family of contrasts. Potthoff (1964) has considered the use of Bonferroni intervals in the context of univariate and multivariate ATI analysis. Values of C appropriate for these intervals are given for some families of interest in Table 5 which will be discussed later. For a given family of interest, each interval should be computed with the approach, Roy-Bose or Bonferroni providing the smallest width. Any example contrasts in the current example are assumed to be members of families including contrasts for various post hoc combinations of dependent variables; therefore, the Roy-Bose intervals will be used. The resulting intervals are larger in most cases than the corresponding Bonferroni intervals for just the original p dependent variables.

One global hypothesis of interest in the current example focuses on the ATI parameters; i.e., $H_0: \beta_5' = \beta_6' = 0'$. As pointed out above, when this H_0 is true, all regression lines are parallel for each dependent variable. This global ATI H_0 is, of course, analogous to the global interaction H_0 in two-way ANOVA/MANOVA. The λ_1 critical value for $\alpha = .05$ is $\lambda_1^{.05}(2, 1/2, 201) = .032$ and $C = 3.61$ for the associated interval estimation. Implied hypotheses and contrasts would include those concerned with pairwise slope comparisons among the treatment levels. For example, if S_1 is the

vector of slope coefficients in the i^{th} treatment level associated with the p dependent variables, some multivariate implied hypotheses would be

$H_1: \underline{S}_1' - \underline{S}_3' = \underline{\beta}_5' = \underline{0}'$; $H_2: \underline{S}_2' - \underline{S}_3' = \underline{\beta}_6' = \underline{0}'$; and $H_3: \underline{S}_1' - \underline{S}_2' = \underline{\beta}_5' - \underline{\beta}_6' = \underline{0}'$. Corresponding implied contrasts for linear combinations of the dependent variables would also be of interest, i.e., $\gamma_1 = S_1 - S_3 = \beta_5$, etc. These interaction contrasts are analogous to the "difference of differences" interaction contrasts suggested by various authors as part of the follow-up to a significant interaction in two-way ANOVA (see, e.g., Marascuillo & Levin, 1970).

For the example data, the global ATI null hypothesis is rejected ($\gamma_1 = .041$). The n_{LH}^2 index of strength of association is .044 and associated univariate results are given in Table 2. The multivariate and univariate indices of strength of association are small, suggesting that the interaction is not important. This impression contrasts, however, with the picture of a modest but practically significant ATI resulting from direct inspection of the interaction patterns in Figure 1, suggesting that judgements based on explained variance only may be misleading. The largest increases in explained variability due to the ATI terms are found for the ACHA and INTRST dependent variables. If a priori univariate ATI hypotheses for each dependent variable had been specified, the ATI effect would have been significant at the .05 level for the ACHA and INTRST variables.

Insert Table 2 about here

Results of the follow-up to the rejected global ATI hypothesis considering specific interaction hypotheses and contrasts are shown in Table 3. Of the three possible multivariate pairwise slope comparisons, only the $\underline{S}_1 - \underline{S}_3$ comparison is statistically significant at the .05 level. The associated simultaneous 95% confidence intervals for each of the original dependent variables are relatively wide compared to the magnitudes of point estimates; they are, however, narrow enough in the case of the $\underline{S}_1 - \underline{S}_3$ contrast for the ACHA and INTRST variables so that there is little doubt about the direction of those interactions.

Insert Table 3 about here

Another possible global H_0 for a starting point, which might be called the "simple treatment effect" global hypothesis, is $H_0: \underline{\beta}_3' = \underline{\beta}_4' = \underline{\beta}_5' = \underline{\beta}_6' = 0$. Note in equation 2 that if the H_0 is true, then all regression lines are coincident for each dependent variable; that is, there is no treatment effect at any level of aptitude. The critical value for the example is $\lambda_1^{.05}(4, -1/2, 201) = .045$, and $C = 4.28$. Possible implied hypotheses would include the multivariate global ATI and slope comparison hypotheses discussed above plus a class of hypotheses representing a multivariate generalization of the Johnson-Neyman technique for identification of regions of significance (Johnson & Fay, 1950). That is,

$$H_1: \begin{bmatrix} \underline{\mu}_1' - \underline{\mu}_3' \\ \underline{\mu}_2' - \underline{\mu}_3' \end{bmatrix} = \begin{bmatrix} \underline{\beta}_3' + \underline{\beta}_5' A_1 \\ \underline{\beta}_4' + \underline{\beta}_6' A_1 \end{bmatrix} = 0$$

could be tested for various A_1 values, resulting in a plot of λ_1 versus A .

Comparison with λ_1^α for the global H_0 would then allow identification of the

global significance region. Similar regions could be identified for one-degree-of-freedom hypotheses, e.g.,

$$H_1: \mu_1' - \mu_2' = (\beta_3' - \beta_4') + (\beta_5' - \beta_6') A_1 .$$

The family-wise error rate is held to .05 for all such hypotheses implied by the global simple treatment effect hypothesis above. Corresponding implied simple treatment contrasts for linear combinations of the dependent variables of interest would be $\gamma = \mu_1 - \mu_3 = \beta_3 + \beta_5 A_1$, etc. Again, regions of significance could be obtained for each univariate contrast. The simple treatment effect global hypothesis and associated follow-up is analogous to the "simple effects design" and follow-up proposed by Marascuilo and Levin (1970) for two-way ANOVA designs, and the two-way ANOVA approach to ATI research suggested by Betz and Gabriel (1978).

The global simple treatment effect hypothesis ($H_0: \beta_3' \dots = \beta_6' = 0'$) is rejected ($\lambda_1 = .225$) for the current example. Note that if the global ATI null hypothesis considered above was included as part of the follow-up to this more general hypothesis, it would not be rejected at the .05 level. Variation of λ_1 values for the simple treatment effect at specific aptitude values ($H_0: \mu_1' - \mu_3' = \mu_2' - \mu_3' = 0'$) is shown in Figure 2. Comparison with the $\lambda_1^{.05}$ of .045 indicates that the corresponding multivariate Johnson-Neyman region of significance extends across the entire aptitude range. As mentioned above, significance regions can also be obtained for specific simple treatment contrast hypotheses. For example, variation of λ_1 for $H_0: \mu_2' - \mu_3' = 0'$ is also shown in Figure 2. Comparison with the same critical value indicates that the multivariate effect is nonsignificant across the entire ability range.

Insert Figure 2 about here

The description of the effect of the treatment as a function of aptitude would be completed with confidence intervals on all contrasts and linear combinations of dependent variables of interest. For example, 95% confidence bands on the $\mu_2 - \mu_3$ contrast for the ACHC variable are shown in Figure 3. The cost of the various dimensions of simultaneous protection can be illustrated for this contrast. First, the narrowest band in Figure 3 is surprisingly wide (considering the sample size of this study), even though it represents simple intervals ($C = t^{\alpha/2}$) with no simultaneous protection at all. The relative lack of estimation precision reflects relatively large standard errors for interaction parameters compared to those for main effect parameters. Simultaneous protection over all aptitude values increases the band width by a factor of 1.25, further protection for all treatment contrasts increases the width by a factor of 1.58 (compared to the simple band), and, finally, further protection for all dependent variables and linear combinations of dependent variables increases the width by a factor of 2.18 resulting in the band labeled "simultaneous" in Figure 3. In this case, the existence of a region of significance depends on the scope of simultaneous protection selected. Even for the widest band, however, the location of the band relative to the zero point still suggests some confidence in the direction of the effect for aptitudes in the range of 50 - 70.

Insert Figure 3 about here

Finally, two more possible global null hypotheses would be $H_0: \beta_2' = \dots = \beta_6' = 0'$ with a critical value of $\lambda_1^{.05}(4, 0, 201) = .049$ and $H_0: \beta_1' = \dots = \beta_6' = 0'$ with $\lambda_1^{.05}(4, 1/2, 201) = .055$. The implied family associated with the first hypothesis would include all hypotheses considered above (e.g., interaction effect and simple treatment effect hypotheses) plus simple aptitude effect hypotheses, i.e., $H_0: \beta_1' = \beta_2' + \beta_5' = 0'$, etc.

This hypothesis is analogous to the "Cell Means Model" suggested by Marascuilo and Levin (1976) for two-way ANOVA and a corresponding multivariate hypothesis used by Betz (1976) to illustrate interaction analysis in two-way MANOVA designs. The family for the second hypothesis would also include predicted value hypotheses, i.e., $\mu_1' = \mu_{10}$. Computation of confidence intervals for μ 's associated with the original dependent variables would result in simultaneous confidence bands on all of the regression lines in Figure 1.

Of course, any planned or a priori hypotheses can be tested with test-wise error rate or with simultaneous protection using the Bonferroni Inequality, regardless of the results of global tests.

Description with Discriminant Functions

A second advantage of a multivariate analysis is the potential for simplification of the ATI description. That is, it may be possible to describe the ATI effect associated with a large number of dependent variables more parsimoniously by the identification and labeling of several dependent variable factors which "capture" the same relationship. This can be attempted with generalized discriminant analysis (Bargmann, 1970) which determines a optimal linear combinations of the dependent variables (discriminant functions)

associated with a given null hypothesis. The first discriminant function is the combination which maximizes the effect represented by the hypothesis, the second function maximizes the effect subject to the restriction of being uncorrelated with the first function, etc. The coefficient vectors for the discriminant functions are the eigenvectors associated with the eigenvalues (λ 's) of $H E^{-1}$ on which the various hypothesis test criteria are based. The λ 's can be used further with several test criteria (Timm, 1975, p. 380) to determine how many of the discriminant functions capture a statistically significant portion of the overall effect. The relative importance of each of the statistically significant functions can be assessed by noting that λ_1 is defined as the ratio of hypothesis variability (sums of squares) to error variability for the 1th optimal combination of dependent variables; that is, $\lambda_1 = \Delta R_1^2 / (1 - R_{F1}^2)$. Correlations between each function and the dependent variables (i.e., loadings) are used in the attempt to label the functions, and rotation of the functions may be used to attempt to improve interpretations (Hall, 1969).

A discriminant analysis can be conducted for any global or implied multivariate hypothesis. Of course, the resulting functions will in general be different for each hypothesis. In addition to obtaining a more parsimonious description for each hypothesis, it would also be hoped that certain meaningful functions would consistently emerge across various global and implied hypotheses of interest, allowing even more parsimony of description.

The global ATI hypothesis for the current example results in two discriminant functions, of which only the first is statistically significant. The loadings for the first function of the global ATI hypothesis are shown

in Table 4, along with loadings for the functions associated with the three implied interaction hypotheses. As discussed above, the $S_1' - S_3'$ contrast is the dominant contributor to the global ATI, so it is not surprising that the loadings are similar for the corresponding functions. The ACHA dependent variable would be dominant and INTRST would play a minor role in the interpretation and labeling. The discriminant functions are different across the interaction contrasts and offer no surprises. For example, the INTRST variable is dominant for the $S_2' - S_3'$ contrast, supporting the impression that would result from inspection of Figure 1.

Insert Table 4 about here

The variation of the discriminant function loadings across the range of ability is shown in Figure 3 for the simple treatment effect contrast $\mu_2' - \mu_3'$. The ACHC and ACHA variables are dominant over most of the ability range to a value of approximately 70 - 75, beyond which the INTRST and NCRSE variables become more important. Thus, an attempt to produce a more parsimonious description of the MT vs. HT contrast might talk of the effect in terms of an achievement factor for students with low to average ability and in terms of an attitudinal factor for students with very high abilities. Again, as with the interaction contrast functions, the impression from Figure 3 does not differ much from that which would be based on the interaction patterns for each dependent variable in Figure 1.

Insert Figure 3 about here

Discussion

It is clear that the nature of an ATI description based on post hoc inferential analysis can change drastically depending on the choice of global null hypotheses. For example, the multivariate global ATI was statistically significant when tested by itself in the current analysis, but nonsignificant when tested as part of the follow-up to a significant global simple treatment effect. Furthermore, the widths of intervals and the associated picture of estimation precision depend on the scope of the families to which the intervals belong (i.e., the scope of the global hypotheses which imply the contrasts of interest). Three approaches to the selection of appropriate global hypotheses or families are considered now. The first two represent the two positions in the ongoing Type IV error debate (Games, 1973, 1978; Levin & Marascuilo, 1972, 1973; and Marascuilo & Levin, 1970, 1976).

One approach, characterized by statistical consistency and elegance, is based on the a priori selection of a small number of logically independent global hypotheses, tests of one or more of the global hypotheses, and the follow-up of each rejected null hypothesis with simultaneous tests and intervals for only those tests and contrasts implied by that global hypothesis. Any test of a hypothesis not implied by the rejected global hypothesis is called a Type IV error by Marascuilo and Levin (1970). The most elegant version of this approach would consist of the identification of a single global null hypothesis which implies all of the tests and contrasts of interest in the study, thus controlling the experiment-wise error rate and confidence level. Proposals and illustrations by Betz (1976), Betz and

Gabriel (1978), Levin and Marascuilo (1972, 1973), and Marascuilo and Levin (1970, 1976) are in this spirit. If it is assumed in the current ATI example that only interaction and simple treatment effect hypotheses are of interest, then the global simple treatment hypothesis would be the starting point for this approach and all follow-up would use the global critical value.

The traditional approach to the analysis of interactions in two-way ANOVA is typical of a more pragmatic approach which is less concerned about statistical consistency. Rejection of the ANOVA interaction hypothesis is commonly followed by tests or intervals on simple main effect contrasts (e.g., Kirk, 1968, p. 188), contrasts which are not implied by the interaction hypothesis. Marascuilo and Levin call this practice a Type IV error, but Games (1973, 1978) defends the procedure as legitimate because it addresses questions of practical importance. This approach in the current ATI example would consist of a test of the global ATI hypothesis, followed by interval estimation for interaction contrasts (using the critical value for $H_0: \beta_5' = \beta_6' = 0'$) and simple treatment effect contrasts (using the critical value for $H_0: \beta_3' = \dots = \beta_6' = 0'$).

Even though interval estimation can be used in each of the two approaches above, they are basically hypothesis testing procedures in the sense that follow-up description is usually only attempted when the associated global null hypothesis has been rejected. Cronbach (1975) has suggested that this approach is very wasteful in ATI research because of the low power for

interaction hypotheses found even in moderately large studies. He has proposed that the null hypothesis be "exorcised" and attention instead directed to description of the ATI with point and interval estimation. The question of the extent of simultaneous protection desired in such a descriptive approach might be answered as follows: identify each aspect of the final description which is based on the simultaneous consideration of multiple contrasts, delineate precisely the associated family, determine the minimum number of linear combinations of model parameters and the minimum number of linear combinations of the dependent variables which would imply the family, and compute and use the associated critical value in the simultaneous interval estimates for that aspect of the description. This approach is in the spirit of the recommendations by Potthoff (1964) who considered interval estimation for simple treatment effect contrasts in multivariate research. Scheffé and Bonferroni critical values proposed by Potthoff are shown in Table 5, along with Roy-Bose critical values for the multivariate case. Critical values are given for the somewhat more general situation in which there are n_A aptitudes and n_T levels in a single treatment factor. The structure of the table reflects the fact that the three "dimensions" in the scope of a family of simple treatment effect contrasts are numbers of aptitudes, treatment contrasts, and dependent variables; family scope is then specified by deciding whether a single "case" or all cases is desired for each dimension. All critical values will fall between the two limiting cases of $t^{\alpha/2}$ for a simple interval and the value for a family of all simple treatment contrasts for all values of aptitudes for all dependent variables. Most cells would have more than one possible critical value; the smallest value for a given family would be used.

The potential value of the descriptive payoff of a multivariate analysis, i.e., the discriminant functions which may lead to a more parsimonious description of effects of interest, should increase as the number of dependent variables (p) and sample size increase. For small p , the description is already relatively parsimonious, and it is unlikely that linear combinations of a small number of variables would yield meaningful labels. Perhaps the greatest potential would be found when ATI research is conducted in the spirit of an "external" factor analysis. That is, the researcher would define an outcome domain of interest, a relatively large number of variables would be selected to represent the domain, and a multivariate ATI analysis would be conducted to identify the set of factors (discriminant functions) which maximize the ATI effect. Large sample sizes would be important for such studies because of the extent of capitalization on chance possible in multivariate analysis (Barcikowski and Stevens, 1975), and also to attempt to keep simultaneous interval widths within reason. Cross-validation procedures could be used to assess the degree of confidence to be placed in factors from one study. Of course, the desire for large samples for both inferential and descriptive reasons must be tempered by recognition of the importance of multiple study replications to identify generalizable ATI's (Cronbach, 1975) and the probability that the class or a larger aggregate may be the appropriate unit of analysis for many ATI analyses (Cronbach, 1976).

References

- Barcikowski, R. S., & Stevens, J. P. A Monte Carlo study of the stability of canonical correlations, canonical weights, and canonical variate-variable correlations. Multivariate Behavioral Research, 1975, 10, 353-364.
- Bargmann, R. E. Interpretation and use of a generalized discriminant function. In R. C. Bose, et al. (Eds.), S. N. Roy Memorial Volume, Essays in Probability and Statistics. Chapel Hill, North Carolina: University of North Carolina, 1970, 35-60.
- Betz, M. A. Interaction analysis in MANOVA. Paper presented at the annual meeting of the American Educational Research Association, New York, April, 1976.
- Betz, M. A., & Gabriel, K. R. Type IV errors and analysis of simple effects. Journal of Educational Statistics, 1978, 3, 121-143.
- Burkman, E. Unhooking high school science: The ISIS Project. The Science Teacher, 1974, 41, 30-32.
- Cronbach, L. J. Beyond the two disciplines of scientific psychology. American Psychologist, 1975, 30, 116-127.
- Cronbach, L. J. with the assistance of Dekker, J., & Webb, N. Research on classrooms and schools: Formulation of questions, design, and analysis. Occasional paper of the Stanford Evaluation Consortium. Stanford, California: Stanford University School of Education, 1976.
- Gabriel, K. R. Simultaneous test procedures in multivariate analysis of variance. Biometrika, 1968, 55, 489-504.

- Games, P. A. Type IV errors revisited. Psychological Bulletin, 1973, 80, 304-307.
- Games, P. A. Nesting, crossing, Type IV errors, and the role of statistical models. American Educational Research Journal, 1978, 15, 253-258.
- Graybill, F. A. Theory and application of the linear model. North Scituate, Mass.: Duxbury Press, 1976.
- Hall, C. E. Rotation of canonical variates in multivariate analysis of variance. Journal of Experimental Education, 1969, 38, 31-38.
- Huberty, C. J. Multivariate indices of strength of association. Multivariate Behavioral Research, 1972, 7, 523-526.
- Johnson, P. O., & Fay, L. C. The Johnson-Neyman technique, its theory and application. Psychometrika, 1950, 15, 349-367.
- Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, California: Brooks/Cole, 1968.
- Levin, J. R., & Marascuilo, L. A. Type IV errors and interactions. Psychological Bulletin, 1972, 78, 368-374.
- Levin, J. R., & Marascuilo, L. A. Type IV errors and Games. Psychological Bulletin, 1973, 80, 308-309.
- Lo, M. Y. Statistical analysis of interaction and its application to data from the Cooperative Research Program in primary research instruction. (Doctoral dissertation, State University of New York at Buffalo, 1973.) Dissertation Abstracts International, 1973, 34, 3146-A. (University Microfilms No. 73-29, 111)
- Marascuilo, L. A., & Levin, J. R. Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of Type IV errors. American Educational Research Journal, 1970, 7, 397-424.

- Marascuillo, L. A., & Levin, J. R. The simultaneous investigation of interaction and nested hypotheses in two-factor analysis of variance designs. American Educational Research Journal, 1976, 13, 61-65.
- Miller, R. A. Simultaneous statistical inference. New York: McGraw-Hill, 1966.
- Peterson, P. L. Interactive effects of student anxiety, achievement orientation, and teacher behavior on student achievement and attitude. Journal of Educational Psychology, 1977, 69, 779-792.
- Potthoff, R. F. On the Johnson-Neyman technique and some extensions thereof. Psychometrika, 1964, 29, 241-256.
- Roy, S. N., & Bose, R. C. Simultaneous confidence interval estimation. Annals of Mathematical Statistics, 1953, 24, 513-536.
- Stevens, J. P. Global measures of association in multivariate analysis of variance. Multivariate Behavioral Research, 1972, 7, 373-378.
- Timm, N. H. Multivariate analysis with applications in education and psychology. Monterey, Calif.: Brooks/Cole Publishing Co., 1975.
- Timm, N. H., & Carlson, J. E. Multivariate analysis of nonorthogonal experimental designs using a multivariate full rank model. Paper presented at the annual meeting of the American Statistical Association, New York City, December, 1973.

Table 1

Model Parameters: Point Estimates and Standard Errors¹

Model Parameter	Dependent Variable			
	ACHC	ACHA	NCRSE	INTRST
β_1	29.05 (4.12)	14.77 (4.38)	.779 (.299)	1.74 (.252)
β_2	.510 (.068)	.584 (.072)	.013 (.005)	.016 (.004)
β_3	21.05 (5.97)	32.80 (6.36)	1.01 (.434)	1.23 (.366)
β_4	10.01 (6.17)	8.79 (6.57)	.466 (.449)	.581 (.378)
β_5	-.176 (.089)	-.309 (.094)	-.009 (.006)	-.015 (.005)
β_6	-.082 (.094)	-.096 (.100)	-.009 (.007)	-.011 (.006)
σ	12.26	13.05	.892	.751

¹Standard error in parentheses.

Table 2

ATI Global Hypothesis: Univariate Results¹

Dependent Variable	R_R^2	ΔR^2	$\Delta R^2 / (1 - R_F^2)$	F	p
ACHC	.375	.006	.010	1.95	.143
ACHA	.398	.017	.029	5.93	.003
NCRSE	.093	.005	.006	1.18	.307
INTRST	.082	.017	.019	3.81	.023

¹ R_R^2 is the coefficient of determination for the restricted or main effects model, R_F^2 is for the full or interactive model, and $\Delta R^2 = R_F^2 - R_R^2$. The degrees of freedom for the F statistic are 2 and 407.

Table 3
Interaction Contrasts: Tests and Intervals¹

Contrast (γ)	λ_1 for $H_0: \gamma = 0$	n_{LH}^2	95% CI(γ)			
			ACHC	ACHA	NCRSE	INTRST
$S_1 - S_3$	.040*	.040	-.17 $\pm$.32	-.31 $\pm$.34	-.01 $\pm$.02	-.02 $\pm$.02
$S_2 - S_3$	.011	.011	-.08 $\pm$.34	-.10 $\pm$.36	-.01 $\pm$.03	-.01 $\pm$.02
$S_1 - S_2$	.015	.015	-.09 $\pm$.31	-.21 $\pm$.33	.00 $\pm$.02	-.01 $\pm$.02

¹The critical value for each multivariate hypothesis is $\lambda_1^{.05}(2, 1/2, 201)$
= .032, n_{LH}^2 is an index of multivariate association, and the intervals are
Roy-Bose intervals associated with the global ATI H_0 .

*Significant at $\alpha = .05$.

Table 4
Discriminant Functions for ATI Effects¹

Dependent Variable	ATI Effect			
	Global ²	$S_1' - S_3'$	$S_2' - S_3'$	$S_1' - S_2'$
ACHC	.476	.475	.399	.443
ACHA	.818	.799	.463	.939
NCRSE	.325	.348	.593	.044
INTRST	.631	.659	.912	.279

¹The table entry is the correlation between the variable and the discriminant function.

²Only the first of the two functions is given for the global ATI test.

Table 5

Critical Values (C) for Simple Treatment Effect Intervals¹

Number of Contrasts ²	df _h ³	Number of dependent variables ⁴	
		One	All
One aptitude value			
One	p	t ^{α/2}	$\sqrt{\frac{df_e F^{\alpha}(p, df_e - p + 1)}{(df_e - p + 1)}}$ t ^{α/2p}
All	n _T -1	$\sqrt{\frac{df_h F^{\alpha}}{t^{\alpha/2g}}}$	$\sqrt{\frac{df_e \lambda_1^{\alpha}}{df_h F^{\alpha/p}}}$ t ^{α/2gp}
All aptitude values			
One	n _A +1	$\sqrt{df_h F^{\alpha}}$	$\sqrt{\frac{df_e \lambda_1^{\alpha}}{df_h F^{\alpha/p}}}$
All	(n _T -1)(n _A +1)	$\sqrt{df_h F^{\alpha}}$	$\sqrt{\frac{df_e \lambda_1^{\alpha}}{df_h F^{\alpha/p}}}$
	n _A +1	$\sqrt{df_h F^{\alpha/g}}$	$\sqrt{\frac{df_h F^{\alpha/gp}}{df_h F^{\alpha/gp}}}$

¹Table entries are C values required for the $CI(\gamma) = \hat{\gamma} \pm CS_{\hat{\gamma}}$ computation.

All entries except the Roy-Bose critical values are those given in Potthoff (1964).

²Bonferroni critical values are based on a family of g treatment effect contrasts.

³Hypothesis degrees of freedom (df_h) for each F statistic across a row are specified in this column. The error degrees of freedom (df_e) is $N - q$.

⁴Bonferroni critical values are based on a family involving p dependent variables or linear combinations of dependent variables.

Figure Captions

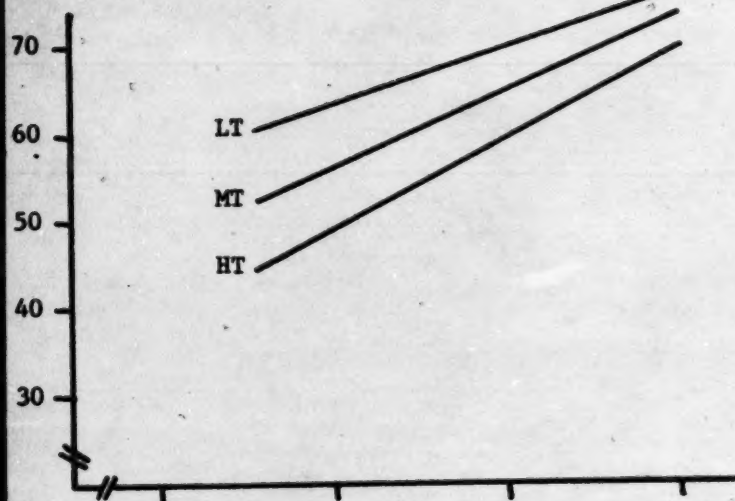
Figure 1. ATI patterns.

Figure 2. Simple treatment effect test statistics.

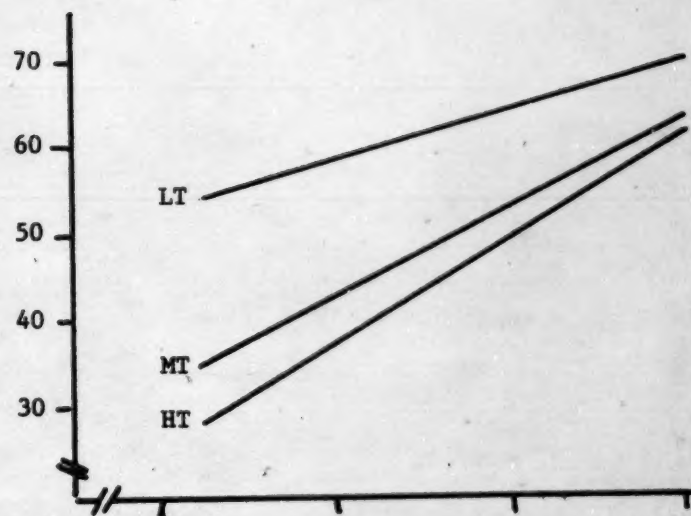
Figure 3. 95% confidence bands on $\mu_2 - \mu_3$ for ACHC.

Figure 4. Simple treatment effect discriminant function loadings.

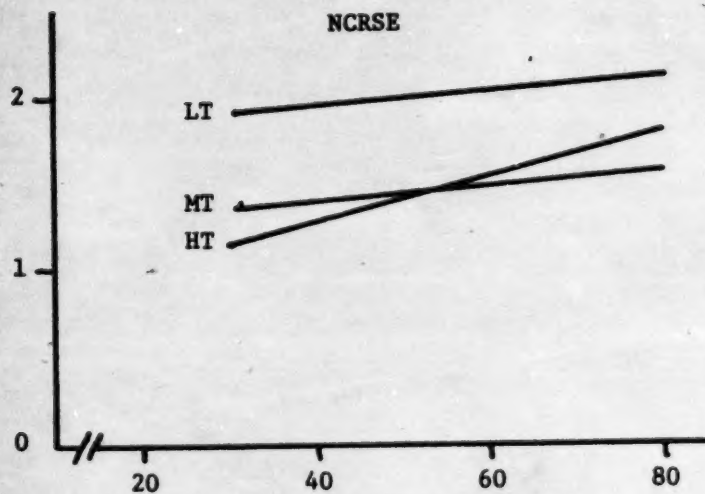
ACHC



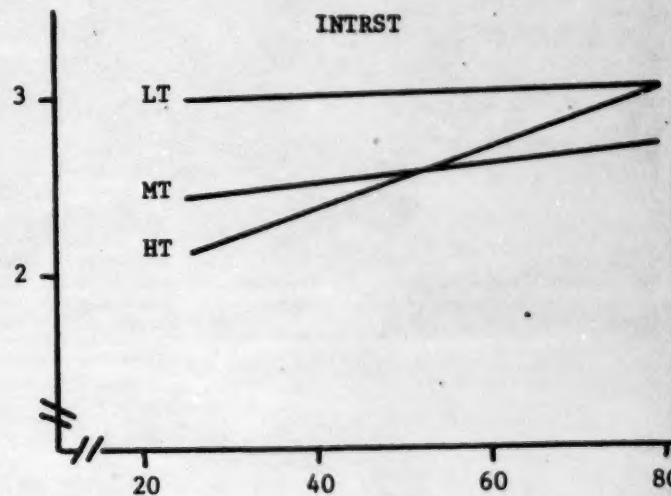
ACHA



NCRSE



INTRST



READING ABILITY

READING ABILITY

